Demystifying Porn 2.0: A look into a major adult video streaming website

Gareth Tyson, Yehia El-khatib, Nishanth Sastry and Steve Uhlig
LOOKING AT PORN
What The Internet Was Made For
Every second $3,075 is spent on porn
Every second $3,075 is spent on porn

By the end of this presentation, you could have earned $3,690,000
12% of the web is porn
42.7% of Internet users view porn
Traffic estimates go as high as 30%
Traffic estimates go as high as 30%

Up to 18% of P2P content
Where is the research?
We want to better understand the nature and characteristics of porn in the Internet
Porn 2.0

- Traditionally porn was static, non-interactive content
  - Bespoke paid websites
- Porn 2.0 has emerged in the last 5 years
  - User video uploads, streaming, downloading, commenting, rating etc.
- Gained huge popularity across the globe
  - There are 5 Porn 2.0 websites in the Alexa Top 100
Porn 2.0: YouPorn

- Prominent Porn 2.0 website
- Founded in 2006
- Boasts of 100 million page views per day
- Several key consumers
  - 19.9% US
  - 9.4% Germany
  - 7.7% Italy
  - 7.4% India
  - 7.3% France
  - 4.1% UK
YouPorn Crawling

• Collected three key datasets from YouPorn
  • Snapshot: A full crawl of all videos
  • 3 day: A repeat crawl three days later
  • Daily: Daily crawls of new videos

• Collected several items of data
  • Number of views, number of ratings, rating, number of comments, duration, category information, upload date, user
YouPorn Crawling

• Collected three key datasets from YouPorn
  • Snapshot: A full crawl of all videos
  • 3 day: A repeat crawl three days later
  • Daily: Daily crawls of new videos

• Collected several items of data
  • Number of views, number of ratings, rating, number of comments, duration, category information, upload date, user

<table>
<thead>
<tr>
<th>Name</th>
<th>Period</th>
<th># Vids</th>
<th># Views</th>
</tr>
</thead>
<tbody>
<tr>
<td>Snapshot</td>
<td>28/02/13</td>
<td>183k</td>
<td>61b</td>
</tr>
<tr>
<td>3 Day</td>
<td>3/03/13</td>
<td>183k</td>
<td>111m</td>
</tr>
<tr>
<td>Daily</td>
<td>1/03 – 4/05/2013</td>
<td>1656</td>
<td>96m</td>
</tr>
</tbody>
</table>
Characterising the content

- Many 'snippets' of content
  - Missing preambles
  - Rarely involve a story or conversation
- Short videos
  - Typically around 5 minutes long
  - 80% below 15 minutes
- This is user (generated) content
- Interesting thought experiment...
  - Usage patterns totally different to most other video services
Injection Rates

78 videos uploaded per day on average

138 recently though

CDF

Video Injection Rate (per day)
Nature of Uploaders

- Not many unique active uploaders: 5,849!
- Highly skewed distribution
  - 56% upload a single video
  - 80% upload at most 5 videos
- The case is the same for amateur content
- Heavy commercial impact
  - Top 100 all commercial players (teasers)
  - Highest non-commercial ranked 180 (116 vids)
- Is this model so revolutionary after all?
Injection Rates

![Graph showing the number of videos injected over time, categorized by whether the uploader is using YouPorn, and whether there is a significant spike in upload activity.]
Injection Rates

75 days without uploads
Upload Processing

• Studied status of each upload (numerical ID)
  • 61% of content is 'being processed'
  • Experimentation showed this was 'being vetted'
  • Effective system
    - Only 11.7% of content is removed
    - Note that removal can only be done by YouPorn

• Means only 18% of uploads become live

• Most true UGC content getting trapped
  • Safer: no copyright risks, all actors over 18, any animals were treated well etc.
The Quirks of Porn 2.0

- Popularity
  - Popularity trends are far flatter in Porn 2.0
  - Users 'spread it around'

- User behaviour
  - Largely ambivalent audience
    - Highly elastic tastes
  - But an impatient one (lots and clicks and skips)

- Product differentiation is extremely difficult
  - And producers often don't particularly try

- Many others...
Any Questions?
Characterising YouPorn's Video Popularity
Head is far less skewed – 65% of views come from top 10%
Popularity Distribution

Tail is more popular – 93% of videos have >10k views.
Number of Views over Time

- Not many accumulated views
Number of Views over Time

Not many accumulated views
Video Popularity over Time

Fraction of Requests vs. Content Age (Days)

- >100k Views
- <100k Views
- <10k Views
Video Popularity over Time

Video popularity plummet over time

Fraction of Requests

Content Age (Days)
Video Popularity over Time

Only very popular content can sustain same level of popularity.
Video Popularity over Time

Fraction of Requests

Content Age (Days)

Still plummets

>100k Views
<100k Views
<10k Views
Video Popularity over Time

Junk videos!
Why?

- Most users browse
  - 63% use front-page browsing
  - 59% use category-page browsing
  - Only 22% visit the site with a particular video in mind
  - Only 9% use links from other sites

- Users are easily satisfied
  - Only 15% find it difficult to locate videos of interest
  - 43% of users are satisfied by over 3/5th of content
Views Per-Page

By far most views

Number of Views (Thousands)

CDF
Views Per-Page

Tiny number of views

CDF

Number of Views (Thousands)
Video Popularity over Time

Fraction of Requests

Content Age (Days)

>100k Views
<100k Views
<10k Views
Video Popularity over Time

Exploitable?
YouPorn's Category System
YouPorn's Category System
Category Density

Strong skew across categories: Top 20% contains 57% of videos
Category Popularity

Uncategorised content: 17% of the corpus vs. 2% of the views
Category (In)efficiency

\[ I = \begin{cases} \frac{V}{C} - 1, & \text{if } V > C, \\ -(\frac{C}{V}) + 1, & \text{otherwise}. \end{cases} \]
Category (In)efficiency 

\[ I = \begin{cases} 
\frac{V}{C} - 1, & \text{if } V > C. \\
-(\frac{C}{V}) + 1, & \text{otherwise.} 
\end{cases} \]
Category (In)efficiency

\[ I = \begin{cases} 
\frac{V}{C} - 1, & \text{if } V > C, \\
-(\frac{C}{V}) + 1, & \text{otherwise.}
\end{cases} \]

Lower density categories are more erratic.
Category (In)efficiency

\[ I = \begin{cases} 
\frac{V}{C} - 1, & \text{if } V > C. \\
-\left(\frac{C}{V}\right) + 1, & \text{otherwise.} 
\end{cases} \]

35 popular

27 unpopular

Inefficiency

Weighted Inefficiency